

Journal policies that encourage data sharing prove extremely effective



*There is currently little incentive for researchers to share their data. But what if it was enough for journals to simply ask authors to make their data available? **Michèle B. Nuijten** reports on a recent study that found journal policies that encourage data sharing to be extremely effective, with a steep increase in the percentage of articles with open data from the moment these policies took effect. Even something as seemingly frivolous as offering a badge to display on your paper as a reward for sharing data can have a transformative effect, not only on the overall availability of data but also on its relevance, usability and completeness, as well as on the rigour and quality of science as a whole.*

For science to work well we should move towards opening it up. That means sharing research plans, materials, code, and raw data. If everything is openly shared, all steps in a study can be checked, replicated, or extended. By sharing everything we let the facts speak for themselves, and that's what science is all about.

Unfortunately, in my own field of psychology, raw data are notoriously hard to come by. Statements in papers such as "all data are available upon request" are [often void](#), and data may get lost if a researcher retires, switches university, or even buys a new computer. We need to somehow incentivise researchers to archive their data online in a stable repository. But how?

Currently it is not in a scientist's interests to put effort into making data and materials available. Scientists are evaluated based on how much they publish and how often they're cited. If they don't receive credit for sharing all details of their work, but instead run the risk colleagues will criticise their choices (or worse: find errors!), why would they do it?

So now for the good news: incentivising researchers to share their data may be a lot easier than it seems. It could be enough for journals to simply ask for it! In our [recent preprint](#), we found journal policies that encourage data sharing are extremely effective. Journals that require data sharing showed a steep increase in the percentage of articles with open data from the moment these policies came into effect.

In our study we looked at five journals. First, we compared two journals in decision making research: *Judgment and Decision Making* (JDM), which started to require data sharing from 2011; and the *Journal of Behavioral Decision Making* (JBDM), which does not require data sharing. Figure 1 shows a rapidly increasing percentage of articles in JDM sharing data (up to 100%!), whereas nothing happens in JBDM. The same pattern holds for psychology articles from open access publisher PLOS (with its data-sharing policy taking effect in 2014) and the open access journal *Frontiers in Psychology* (FP; no such data policy).

Similarly, the journal *Psychological Science* (PS) also contained increasing numbers of articles with open data after it introduced its [Open Practice Badges](#) in 2014. You can earn a badge for sharing data, sharing materials, or preregistering your study. A badge is basically a sticker for good behaviour on your paper. Although this may sound a little kindergarten, believe me: you don't want to be the one without a sticker!

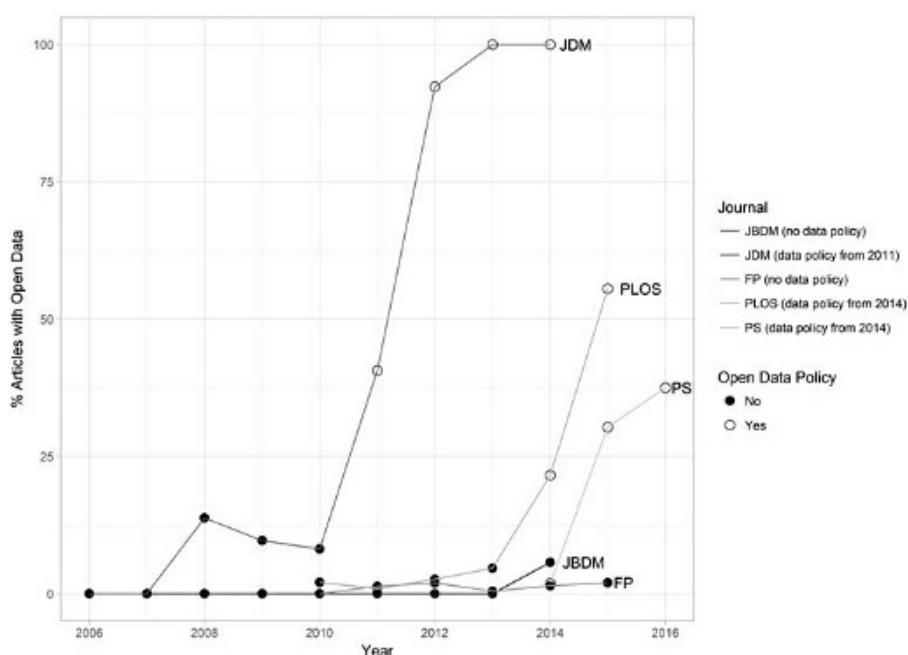


Figure 1: Percentage of articles per journal to have open data. A solid circle indicates no open-data policy; an open circle indicates an open-data policy. Source: Nuijten, M. B., Borghuis, J., Veldkamp, C. L. S., Alvarez, L. D., van Assen, M. A. L. M., & Wicherts, J. M. (2017) “[Journal Data Sharing Policies and Statistical Reporting Inconsistencies in Psychology](#)”, PsyArXiv Preprints. This work is licensed under a [CC0 1.0](#) Universal license.

The increase in articles with available data is encouraging and has important consequences. With raw data we are able to explore different hypotheses from the same dataset, or combine information of similar studies in an Individual Participant Data (IPD) meta-analysis. We could also use the data to check if conclusions are robust to changes in the analyses.

The availability of research data would increase the quality of science as a whole. With raw data we have the possibility to find and correct mistakes. On top of that, the probability of making a mistake is likely to be lower once you have gone to the effort of archiving your data in such a way that another person can understand it. The process of archiving data for future users could also provide a barrier to taking advantage of the [flexibility in data analysis](#) that could lead to false positive results. Enforcing data sharing might even deter fraud.

Of course, data-sharing policy is not a “one-size-fits-all” solution. In some fields of psychological research (e.g. sexology or psychopathology) data can be very personal and sensitive, and can’t simply be posted online. Luckily there are increasingly sophisticated techniques to anonymise data, and often materials and analysis plans can still be shared to increase transparency.

It is also important to acknowledge the time and effort it took to collect the original data. One way to do this is to set a fixed period of time during which only the original researchers have access to the data. That way they get a head start in publishing studies based on the data. When this period is over and others can also use the data, the original authors should, of course, be properly acknowledged through citations, or even, in some cases, co-authorship.

There are many different ways to encourage openness in science. My hope is that more journals will soon follow and start implementing an open-data policy. But aside from merely requiring data sharing, journals should also check if the data is actually available. To illustrate the importance of this, our study found one third of *PLOS* articles claiming to have open data, actually did not deliver (for similar numbers, see the [data by Chris Chambers](#)).

And many (including myself) would even like to go one step further. Datasets should not only be available, they should also be stored in such a way that others can use them (see the [FAIR Data Principles](#)). A good way to influence the usability of open data might be the use of the Open Practice Badges. It turned out that in PS, the badges not only increased the availability of data, but [also the relevance, usability, and completeness of the data](#). Another way of ensuring data quality, but also recognition for your work, is to publish your data in a special data journal, such as the [Journal of Open Psychology Data](#).

Even though data sharing in psychology is not yet the status quo, several journals are already helping our field take a step in the right direction. As a matter of fact, the American Psychological Association (APA) has [recently announced](#) it will give its editors the option of awarding badges. It is very encouraging that journal policies on data sharing, or even an intervention as simple as a badge to reward good practice can cause such a surge in open data. Therefore, I hereby encourage all editors in all fields to start requiring data. And while we're at it, why not ask for research plans, materials, and analysis code too?

I would like to thank Marcel van Assen for his helpful comments while drafting this blog.

This blog post is based on the author's co-written article, "[Journal Data Sharing Policies and Statistical Reporting Inconsistencies in Psychology](#)", a preprint currently available at PsyArXiv Preprints (DOI: [10.17605/OSF.IO/SGBTB](#)).

Featured image credit: [Be free](#) by opensource.com (licensed under a [CC BY-SA 2.0](#) license).

Note: This article gives the views of the author, and not the position of the LSE Impact Blog, nor of the London School of Economics. Please review our [comments policy](#) if you have any concerns on posting a comment below.

About the author

Michèle B. Nuijten is a PhD candidate at the [Meta-Research Center](#) at Tilburg University, The Netherlands. Her research focuses on improving science in psychology and includes topics such as replication, publication bias, and statistical errors. She is also co-developer of "[statcheck](#)", a free and open source tool to check if statistics in articles are internally consistent. Her ORCID iD is [0000-0002-1468-8585](#).